

Criterion 1: Depth of understanding of formation of a hypothesis and coherence of a conclusion

Level of Performance	2-3 (Low)	3-4	4-5 (Medium)	5-6	6-7 (High)
Identifies a hypothesis as an assertion identifying an effect but no cause.	Identifies a hypothesis as an assertion identifying an effect but no cause.	Identifies a hypothesis as an assertion identifying a predictor and a variable.	Formulates a hypothesis that identifies a predictor and a variable.	Formulates a hypothesis that identifies a predictor, testable variables and a variable.	Formulates a well-structured hypothesis that identifies a specific predictor, logical variables and a variable.
Identifies broadly aspects of the scope or constraints of the hypothesis.	Identifies broadly aspects of the scope or constraints of the hypothesis.	Identifies some constraints on the hypothesis and states its scope in terms of what has and has not been covered.	Identifies some constraints on the hypothesis and states its scope in terms of what has and has not been covered.	Identifies some constraints on the hypothesis and states its scope in terms of what has and has not been covered.	Identifies all relevant and specific constraints on the hypothesis and states its scope in terms of what has and has not been covered.
Identifies limited findings derived from the manipulated dataset. Limited interpretation of the manipulated data.	Identifies some patterns and relationships in the manipulated dataset and formulates some findings. Some interpretation of the manipulated data.	Identifies patterns and relationships in the manipulated dataset to formulate a range of findings. Generally accurate interpretation resulting in valid findings.	Identifies accurately a set of findings derived from creating key relationships and patterns in the manipulated dataset.	Identifies insightful and accurate findings based on a rigorous interpretation of the manipulated dataset.	
Makes some observations about the findings and draws a conclusion to the hypothesis. There is little connection between the findings and the conclusion.	Draws a conclusion to the hypothesis based on the manipulation of the findings. There is little connection between the findings and the conclusion.	Interprets findings to identify valid and connections. Draws a conclusion to the hypothesis based on the manipulation however, it cannot be fully substantiated.	Interprets findings to identify key connections and draws from these a valid conclusion to the hypothesis.	Interprets findings to identify key connections and draws from these a valid conclusion to the hypothesis that is fully substantiated.	

### Criterion 1: Hypothesis Formation and Conclusion

#### Hypothesis

- prediction, logical variables, testable,

Students must generate their own hypothesis; it must not be teacher provided.

A short report that will be in two (2) parts:

- Hypothesis and Scope and Constraints
  - Findings and a Conclusion (confirm/refute hypothesis)
- NB: No specific format stated (written, visual or aural)

The findings are based on the interpretation of the manipulated data, and the conclusion is consistent with the findings.

[SD p. 28, 30, 31, 38; Administrative Information p. 2]

### VCAA Definition of a Hypothesis

A hypothesis typically describes a **cause and effect** and should include **at least one prediction** that does not arise from another explanation.

It is a statement or **tentative** prediction about the relationship between **two or more variables**; that a change in one of these variables will result in some change in another.

A hypothesis must be **testable**.

Reasons outlining initial prediction given.

(VCAA Advice to Teachers)

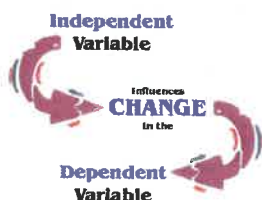
<http://deeddigipubs2014.businesscatalyst.com/vcaa/vce-computing/vce-computing-units-3-and-4-informatics-unit-3-learning>



### Two or More Variables

- Independent variable (IV):** the variable in an experiment that is **changed or manipulated**
- Dependent variable (DV):** the variable in an experiment that **responds to change**.

NB: You can only test one variable at a time.



### Hypothesis Formation #1

Independent Variable  
(what you are changing)

Dependent Variable  
(what will change)

Listening to **music** for more than an hour will lead to lower test scores than studying without music because you are not completely focussed on the topic.

Prediction  
(what will happen)

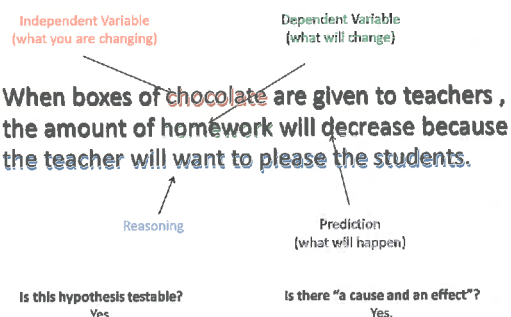
Reasoning

Sample structure:  
Is this hypothesis testable?

The **independent variable** (what you are changing) then **dependent variable** (what will change) including the prediction (what will happen) because **reasoning** (explanation).

Is there "a cause and an effect"?

## Hypothesis Formation #2



## Hypothesis Examples

1. Student numbers studying Physics at Cooke School has increased due to advertising the opportunity to attend the Luna Park Physics Week because students can see the relevance and importance of Physics in everyday life.
2. The number of basketball premierships won by Dunkers College has increased due to the introduction of a three year VCE and 19 year old students are much bigger and stronger and have more skills.
3. The number of people playing a video game increases whenever an expansion pack is released because of the renewed interest in the game.

## Hypothesis Examples

4. The number of lyrebirds in the Dandenong Ranges has increased significantly in the past five years due to the implementation by the Council of measures to control feline population and behaviours.

5. Females playing football has increased in the past five years due to the increased funding and support by the AFL and the publicity that has been generated.

6. The number of hipster cafes in inner Melbourne has increased as the number of people who identify themselves as hipsters have.

## Good or Bad Hypotheses?

	Potential Hypothesis	Good or Bad?	Reason
A	Chocolate ice cream melts faster than vanilla ice cream.		
B	Will fertilizer make my lawn grow more?		
C	Pepsi is better than coke.		
D	Shooting 500 baskets a day will improve my shooting stats.		
E	Will extra water make my plants grow faster?		
F	All hipsters live in Fitzroy.		
G	All hipsters live in Fitzroy.	bad	no cause and effect

## Solution specifications

When creating a solution to a problem we need to consider a number of factors which make of the specifications of the solution.

The specifications of a solution consist of:

- ▶ Requirements
- ▶ Constraints
- ▶ Scope

## Requirements

What your information needs to do, the qualities it should have and what data is required to create the information.

For example think about the output required (info) by the users. What should it tell them? How are they going to use the info?

When this has been decided then the required data can be identified.

## Requirements

**Functional requirements:** the tasks that the solution should be able to perform, eg: what reports, charts etc are required by the user.

SAT Outcome, that it lets your reach a valid & substantiated conclusion as to whether your hypothesis is supported or refuted

**Non-functional requirements:** the attributes or qualities of the info, eg: accurate, easy to use, secure etc.

**Data requirements:** the data selected must be able to produce the required output.

## Constraints

The limits and restrictions under which your solution must be produced.

This will usually reduce your freedom of design choice.

## Constraints

Constraints will usually fall into five categories:

- ▶ **Economic:** relating to cost & time. How long can you afford to spend on creating the solution?
- ▶ **Technical:** availability of equipment (hardware & software), capacity limitations and security requirements.
- ▶ **Social:** expertise of users.
- ▶ **Legal:** legislation such as copyright & privacy.
- ▶ **Usability:** can intended users operate the solution. Does it serve its intended purpose?
- ▶ Example, could not get permission, de-identify the respondents, ran out of time, could not get a picture you wanted

## Scope

States the boundaries or parameters, What the info produced must achieve and what it is not required to achieve.

For example what info does your solution need to produce to support or refute your hypothesis; what will or will not be covered

This should be defined precisely so developers can effectively allocate time & resources to the project.

Scope largely defined by its functional & non-functional requirements



### What is data?

Data is made up of raw, unprocessed facts & figures.

At the data stage it is not in a useful form capable of solving the solution.

When it has been converted (processed) into a useful form then it can be used (for example to support or refute a hypothesis).

### Categories of data

Can be categorised in a number of ways, however the most common classifications are:

- ▶ Primary & secondary
- ▶ Quantitative & qualitative

### Primary data

Has not been filtered by interpretation or evaluation (unprocessed state).

Usually collected directly by the user however old data that has never been interpreted can be primary data.

Strengths: specific to needs, can be trusted because know source.

Weaknesses: cost & time.

### Secondary data

Collected and **interpreted** by someone else.

Examples include professional researchers & Govt. organisations.

Strengths: cheap & quick to collect, huge datasets available, can be useful to support your own findings.

Weaknesses: not specific, sources may not be reliable, may not be able to access original unprocessed data.

### Quantitative data

Concerned with numbers and measurement.

It uses an objective approach (no opinions) & closed questions (limits the response).

Strengths: simple to collect, store and analyse. Can easily be compared to previous results.

Weaknesses: limits the responses and does not give people the chance to give their opinion.

### Qualitative data

Expressed in words.

It uses a subjective approach (opinions, personal views & experiences), open questions (allows a range of responses).

Strengths: can be used to gain a wider range of responses. Allows people to elaborate.

Weaknesses: difficult to collect, store and analyse (needs to be encoded to be analysed statistically).

### **Coding qualitative data**

Because it can be difficult to work with it is often simpler to encode the data.

This requires responses to be categorised and then allocated a value (see p 86).

Coding reduces original wordiness using freely chosen summary terms, descriptive coding

Rubrics, a detailed list of descriptive grading criteria that correspond with a code, mark)

Some software tools are available to do this

### Data sources

- ▶ Treat data from unidentified sources with care and take steps to authenticate the data (second-sourcing).
- ▶ Establish a “chain of custody” – from creator to you.
- ▶ A reputable data source that has authority will more likely provide high-quality data.

### Data sources

Because there is so much data available (especially online) it is vital the data selected is appropriate.

A combination of primary & secondary sources can assist in having the data available to solve an information problem (such as supporting or refuting a hypothesis).

Data integrity is very important. This will be discussed later.

### Data sources

There are hundreds of data sources available online in Victoria and Australia alone.

- ▶ Libraries, whether local, state or national.
- ▶ Government departments and agencies (BOM, ABS etc.).
- ▶ Private organisations such as businesses or societies.

### Acquiring data

Data can be acquired using a variety of methods including:

- ▶ observation
- ▶ interviews
- ▶ Surveys & questionnaires
- ▶ Querying existing databases

However to acquire data you first need to know how to ask questions.

### Acquiring data Asking questions

#### Closed questions

- ▶ Usually found in surveys & questionnaires.
- ▶ Limited choice of responses, so can use tick boxes, radio buttons etc.
- ▶ Quicker & easier to collect and collate data.

### Acquiring data Asking questions

#### Open questions

- ▶ Usually found in interviews.
- ▶ No limits on responses. More able to give opinions and ask follow-up questions.
- ▶ Responses are more detailed and unique.
- ▶ More time consuming to collect and difficult to collate data.

### Acquiring data Asking questions

#### Faults in questions

Some questions can be flawed which impacts on the quality of the data collected.

- ▶ Loaded questions can disguise a hidden agenda. "Why do you think immigrants are involved in so much crime in Australia".
- ▶ Leading questions can be biased in certain directions. "Do you think murderers should get life imprisonment or the death penalty"?
- ▶ Closed questions can often be biased in this way through the options they allow.

### Acquiring data Interviews

- ▶ Real time, personal interactions.
- ▶ Allows the questioning to be shaped by the responses given.
- ▶ Can pick up on subjects body language.
- ▶ Can be conducted using video technology if required.
- ▶ Record if possible.
- ▶ Important that interviewer is well prepared and that subject is comfortable and relaxed.

### Acquiring data Observations

- ▶ Observing behaviour can be informative because people are not influenced by the data collector (covert v overt).
  - ▶ Can pick up on non-verbal behaviour (eg: body language).
  - ▶ May have electronic aspects of measuring (eg: audit trails and log-ins).
  - ▶ Costly in terms of time & money.
- Susceptible to observer bias.

### Acquiring data Surveys & questionnaires

- ▶ Often used as starting point for research when little is known about the subject.
- ▶ Used with larger groups often in a written (or online) format.
- ▶ A great deal of data can be gathered relatively quickly and cheaply.
- ▶ Don't allow for clarification or follow-up with the subject.
- ▶ Easy for respondents to be untruthful, leave out questions or respond in the way they think the data collector would expect.

### Acquiring data Querying resources

- ▶ Often the data you require will be contained in large, complex data sets, so an ability to query (filter) the data will be vital.
- ▶ Usually built-in services are available to do this.
- ▶ Query by example (QBE) - type in a description of the data you are seeking (eg: Google). See p 105.
- ▶ Structured Query Language (SQL) - common method of requesting data from a database. It actually creates code to search for your query. Your queries in Access were created this way. See p 106.

### Acquiring data Referencing data sources

It is important that any data sources you use are acknowledged in your work. We do this to avoid:

- ▶ Claiming other people's work as our own (plagiarism).
- ▶ Readers can find the original source if required.
- ▶ Copyright laws are observed.
- ▶ Moral rights are observed.



**Acquiring data**  
**Referencing data sources**

- ▶ This includes direct quotes as well as ideas, summaries or paraphrasing.
- ▶ Various methods of acknowledgement can be used which we will investigate later (see p 109).



### **Data integrity**

- ▶ It is vital if we are to produce useful information that the data we use can be trusted (ie: it has integrity).
- ▶ GIGO
- ▶ Whether storing, transmitting or archiving data integrity must be maintained.

### **Data integrity**

Factors that influence data integrity include:

- ▶ Timeliness
- ▶ Authenticity
- ▶ Relevance
- ▶ Accuracy

### **Data integrity**

#### **Timeliness**

- ▶ Should be processed while current and retrieved without significant delays.
- ▶ Decisions should not be based on outdated data.
- ▶ Relies on capabilities of hardware & software used. Chances of potential delays should be reduced.
- ▶ Age of data should always be checked before using.

### **Data integrity**

#### **Authenticity**

- ▶ Will be reliant on the reliability of the source and whether this can be verified.
- ▶ Has the data been corrupted or changed?
- ▶ Can the data be second-sourced or check the original documents.

### **Data integrity**

#### **Relevance**

- ▶ This measures how closely a resource corresponds with the person's desire for information.
- ▶ Reasons that data can lose relevance include relating to a different time or place, differences in circumstances or getting off topic.

### **Data integrity**

#### **Accuracy**

Two main characteristics of accuracy are:

- ▶ Content (functionality). For example correctness & completeness.
- ▶ Form (appearance). For example clarity and consistency (of format).

### **Data integrity**

#### **Accuracy – Correctness**

- ▶ Values are correct.
- ▶ Caused by faulty acquisition methods (including equipment failure, misinterpretation of questions, fraud or vandalism, small sample size, data entry errors and time factors)
- ▶ Remove or repair data from other sources. Use validation techniques.

### **Data integrity**

#### **Accuracy – Completeness**

- ▶ No missing data (can be difficult to achieve).
- ▶ May have been lost, deleted or unable to be retrieved.
- ▶ Relevant questions may not have been asked or all questions have not been answered.
- ▶ Refer to original data if possible. Ensure fields are not left empty (existence validation).

### **Data integrity**

#### **Accuracy – Clarity**

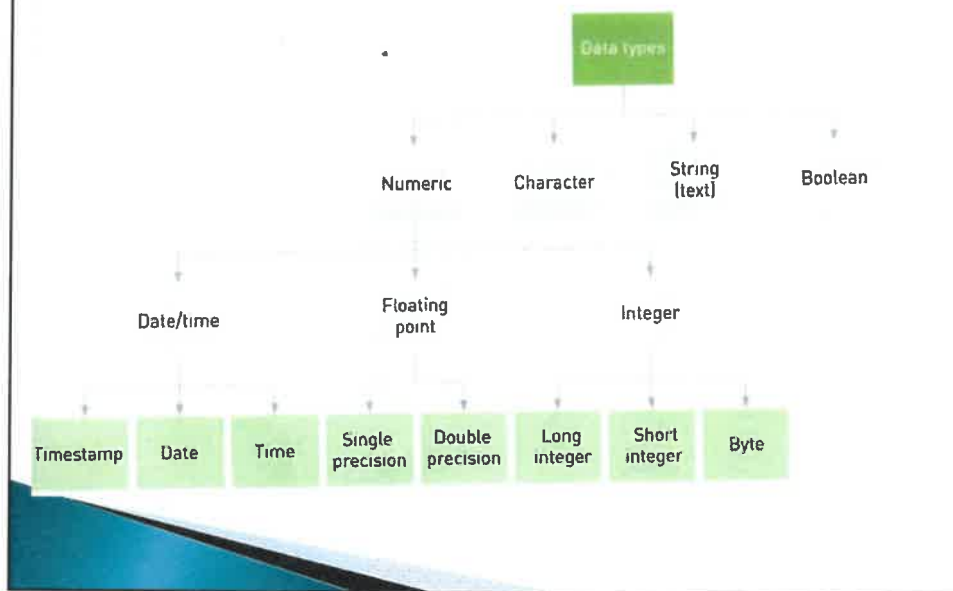
- ▶ Data is clearly presented so it cannot be misinterpreted.
- ▶ Make sure rules for the format of particular data fields (eg: dates).

### **Data integrity**

#### **Accuracy – Consistency**

- ▶ The format of the data must be consistent.
- ▶ Inconsistent use of place names (eg: st. or street) or dates (mm/dd/yyyy) or (dd/mm/yy) can make manipulating the data more difficult.
- ▶ Can also be a problem with discrepancies between multiple data sources (eg: different first names – Matthew or Simon).
- ▶ Use validation and data checking between databases. Also ask the same question in different ways to see if the answers are consistent (student/staff/parent surveys).

## Data types & data structures



## Data types & data structures

- ▶ A Field's data type is separate from its data format
- ▶ A timestamp data type contains both a date and a time of day
- ▶ Integer type cannot store fractional data
- ▶ Spreadsheets guess what data type to use
- ▶ Databases have formal fields, records & tables while spreadsheets do not

